PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 31 October 15, 1945 Number 10

Copyright 1945 by the National Academy of Sciences

STRUCTURE OF THE SALIVARY GLAND CHROMOSOMES OF DIPTERA

By Hans Ris and Helen Crouse

Rockefeller Institute for Medical Research, New York, and Department of Zoölogy, University of Pennsylvania*

Communicated September 4, 1945

The giant chromosomes of the Diptera constitute ideal material for studies on chromosome structure and chemistry and for cytogenetic analysis because of their tremendous size and obvious pattern of longitudinal differentiation. So far it has been impossible to take full advantage of them, however, because they have not been fully understood in terms of the structure of mitotic chromosomes.

In the literature we find three types of interpretation of the banded appearance of the giant chromosomes: (1) The chromosome is composed of several helically coiled threads. The gyres of this coil appear as bands.[1–3] (2) The chromosome is formed of a bundle of completely relaxed chromonemata which originated endomitotically; because of somatic synapsis homologous chromomeres join to form bands.[4–7] (3) The chromosome consists of a large number of chromonemata which are submicroscopic and therefore invisible. The visible structures do not correspond to chromomeres or chromonemata but originate in a different manner.[8]

None of these hypotheses is wholly satisfactory. The first one has been discredited because it is impossible to interpret the bands of the mature giant chromosomes as gyres of a simple large coil. On the other hand, many investigators have seen coils in the giant chromosomes, and bands and coils have even been found to occur simultaneously in the same nucleus.[1, 2, 9–11] A satisfactory interpretation of the giant chromosomes must take this evidence into account.

The second hypothesis has been adequately criticized by Metz. It cannot account for the vesiculated appearance which the giant chromosomes often show, nor for the length of the chromonemata unless some growth of

the chromosome is assumed in addition to uncoiling. This theory, nevertheless, is accepted today by most cytologists because it allows a uniform interpretation of chromosome structure based on the chromomere hypothesis. Recent work on plant and animal chromosomes, however, has shown that the chromonema is uniform and not a series of chromatic granules on an achromatic thread.[12–16] The chromomeres are misinterpretations of coiled structures or points of overlap of chromonemata. Therefore, since mitotic chromosomes are not composed of true chromomeres, it seems doubtful that bands can be interpreted as aggregates of chromomeres.

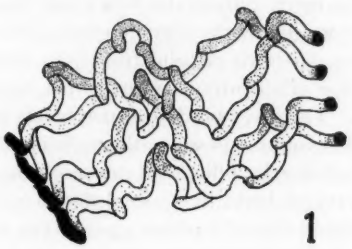
The third hypothesis, assuming submicroscopic chromonemata, places the giant chromosomes in a class by themselves since the component chromonemata are visible in all other types of chromosomes.

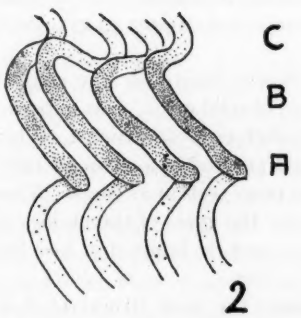
The purpose of this paper is to suggest an interpretation of the structure of giant chromosomes which is in harmony with the latest knowledge of chromosome structure in general (meiotic chromosomes of Tradescantia and the grasshopper, lamp-brush chromosomes of the frog oöcyte[15, 16]) and which can account for the apparently conflicting observations of different investigators. Proof of our hypothesis rests on a detailed study of the development of the giant chromosomes. This work is in progress.

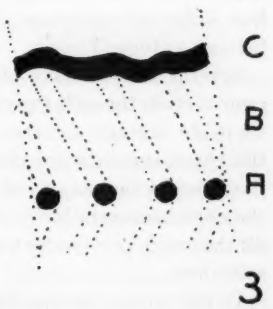
According to our hypothesis the salivary chromosome (Sciara) consists of a definite number of chromonemata which are coiled in a complicated fashion. The coiled threads can be seen most clearly in regions where the banded organization has been disrupted (so-called puffed regions). These regions occur at definite loci along the chromosomes of the larva and become increasingly prominent in the pupa as histolysis sets in. Figure 1 illustrates our interpretation of such a puffed region. Along each chromonema the gyres of a narrowly pitched helix ("minor coil") can be seen. The chromonemata do not run parallel to the length of the chromosome but proceed in an irregular manner, weaving back and forth across the width of the chromosome. Chemical agents which are known to uncoil the mitotic chromosome (KCN, $NaHCO_3$, hot water) can transform normally banded material into a similar mass of threads. These observations suggest that the banded chromosome is composed of coiled chromonemata. It will now be demonstrated how the coiled chromonemata can form bands.

Chromosomes from untreated medium aged Sciara larvae are best suited for this demonstration. At this stage of development four chromonemata can be followed in each homologue. Figure 2 represents our interpretation of two successive bands in terms of coiled chromonemata. It can be seen that the chromonemata are thrown into wide gyres ("major coil") as they proceed along the length of the chromosome. In region *A* they are running vertically, and therefore in optical cross section give the appearance of a granular band (Fig. 3). From here each chromonema can be traced as it

runs horizontally across an interband region *B*. Because of the major coil, the chromonemata of the interband region always run diagonally as most investigators have observed. Only in greatly stretched areas do they run parallel to the long axis of the chromosome. The much lighter appearance of the interband regions we explain as follows: (1) the chromo-







FIGURES 1-3

Diagrammatic representation of the structure of salivary gland chromosomes in a medium-aged larva (Sciara). Only one homologue is drawn.

Fig. 1. Puffed region with bands disrupted.

Fig. 2. The course of the chromonemata through two consecutive bands. (The minor coil is omitted.)

Fig. 3. Appearance of the same region at a medium focal level.

nemata run vertically in the band regions and horizontally in the interbands; consequently, much more light is absorbed in the bands; (2) one usually focuses on a longitudinal optical section of the chromosome, for here the bands are clearest. In such a section, however, more chromo-

nemata are in focus in the band than interband regions; (3) the salivary chromosomes are usually examined in smear preparations. Smearing stretches the chromonemata disproportionately in the interband regions, for the bands, where the chromonemata run transverse to the chromosome axis, resist stretching. In sectioned material stained by the Feulgen reaction we find that chromonemata are much more conspicuous in the interbands. Region *C* in figure 2 illustrates how a solid (i.e., non-granular) band could arise. Here we follow the chromonemata into the next gyre of the major coil. As they dip down, running from left to right across the chromosome, the impression of a continuous line rather than separate granules is produced (Fig. 3). The wavy border of such bands is caused by the minor coil of the chromonemata. Because the individual threads of region *C* proceed at different levels, filling out most of the cross section of the chromosome, this region looks like a solid disc (i.e., threads at all levels). Sometimes in chromosomes of medium aged larvae several gyres of a two-stranded helix can be seen within one homologue.[1-3, 9-11] This appearance can best be understood if we assume that the four chromonemata are closely appressed in pairs, forming the wide gyres of the major coil. When these threads have separated laterally, a banded structure is produced as has been shown above. Such an interpretation makes it possible to understand how different regions on the same chromosome may appear as gyres of a helix or as typical bands.

Several authors have observed that the bands often look like rings in cross section through the chromosome instead of solid discs. Our hypothesis could account for the apparently contradictory observations. When the chromonemata are closely appressed and proceed in a common helix, the gyres of the major coil appear in optical cross section as rings. When the chromonemata have come apart, however, the gyres of the major coil fill the entire cross section of the chromosome, and the bands now look like solid discs.

In old larvae the banded chromosomes have the same structural characteristics, the only difference apparently being the greater number of threads brought about by endomitosis.

If the chromonemata in these giant chromosomes have both minor and major coils, as appears to be the case, they must have grown enormously in length. Such a growth of chromonemata is, however, not restricted to the banded chromosomes. A comparable increase in length occurs, for instance, in the "lamp-brush" chromosomes of certain oöcytes where recent studies[16] have shown that the characteristic loops are actually the gyres of the major coil of the chromonema and contain in addition the minor coil. In contrast to this great increase in length, the chromonema seems to remain approximately constant in diameter. Growth is thus restricted to the long axis of the chromosome. Some authors have suggested that the

longitudinal growth occurs mainly in "nongenic" material of the chromonema. Since giant chromosomes are always found in cells which are especially active physiologically, and since no visible differentiation into alternate "genic" and "inert" regions is visible, it seems to us more likely that the genes themselves increase in mass. Heterochromatic regions which have been shown to contain few genes (Drosophila) do not exhibit in the salivary chromosomes longitudinal growth comparable to that of euchromatic regions. This is in agreement with the view that the genes themselves grow rather than the inert, non-genic material.

The extensive cytogenetic work on Drosophila has shown that the bands in the giant chromosomes can be correlated with definite gene loci. If the bands are due to specific coiling, an interesting relationship between the coiling and the specific molecular structure of the chromonema becomes apparent. This complex coiling pattern of the giant chromosomes could be an expression of the longitudinal differentiation in the gene string. The detailed correspondence of coiling and gene specificity poses an interesting problem for investigation.

Cooper[17] has pointed out that the increase in number of strands in a helix causes it to uncoil. This, however, is true only as long as the component threads remain closely appressed in a common helix. It no longer holds if the chromonemata can separate laterally, as appears to be the case in the giant chromosomes.

It is generally believed that somatic as well as meiotic synapsis is the consequence of a complete uncoiling of the chromosomes. If this were true, it would constitute a serious objection to our assumption that the somatically synapsed giant chromosomes consist of coiled chromonemata. However, the demonstration of a typical helix in synapsing meiotic chromosomes of Tradescantia and the grasshopper[15, 16] invalidates this objection. The chromosomes are coiled when they synapse in meiosis, and they are coiled when they synapse somatically in young dipteran larvae.

It is generally assumed that the chromonemata within one chromosome are held together by the same forces that result in synapsis of homologues. This assumption is unjustified. Rather, it appears that the chromonemata within a chromosome are held together in the same fashion as the coiled chromatids of ordinary prophase chromosomes. It has been suggested[18] that achromatic material binds the chromatids together. Experiments with hypo- and hypertonic solutions on salivary chromosomes[19] seem to confirm this hypothesis. In hypotonic solutions the chromonemata separate, and bands disappear. This can be reversed in isotonic medium. With hypertonic Ringer the chromosome shrinks in diameter, and the bands appear more distinct. All these facts can best be explained by a reversible swelling and shrinkage of an achromatic substance between the coiled chromonemata.

The most puzzling appearance of the salivary chromosome is the so-called vesiculated condition. This can occur at localized regions in different developmental stages of the Sciara larva. In certain physiological states all the giant chromosomes of larvae in the same culture may be vesiculated. Under other conditions the banding pattern completely disappears and the chromosomes look like masses of faintly staining coiled threads (ghost chromosomes). Appearances similar to these can be induced experimentally in normally banded material. Immersion for a few seconds in 1 *M* NaCl causes the chromosomes to become vesiculated. Prolonged treatment with the salt solution produces typical ghost chromosomes. We interpret these phenomena in the following manner. Around each chromonema there is a Feulgen-positive substance, which, under certain conditions, can come off the threads and form chromatic connections between chromonemata and between chromosomes (chromatic coating[18]). Such connections between the coiled chromonemata of the giant chromosomes give the appearance of vesiculation. As was shown by Mirsky and Pollister,[20] nucleoproteins can be dissolved from chromosomes with 1 *M* NaCl. The vesicles observed in the giant chromosomes after short treatment with the salt solution we believe to be caused by the initial dissolution of this substance. Prolonged treatment completely dissolves these nucleoproteins; ghost chromosomes and a basophilic cytoplasm result. This basophily of the cytoplasm is likewise characteristic of untreated ghost cultures.

Summary.—On the interpretation presented here the giant chromosomes of dipteran larvae consist, like mitotic chromosomes, of a number of helically coiled chromonemata. The chromomeres in the giant chromosomes are misinterpretations of coiled structures, as has been demonstrated for meiotic and mitotic chromosomes. It is shown here how the appearance of bands and interbands may be caused by complex coiling of a bundle of chromonemata. The chromonema itself is uniformly Feulgen-positive. The giant size of these chromosomes would then be due to: (1) great increase in length of the chromonema (longitudinal growth of individual genes); (2) increase in the number of chromonemata by endomitosis; (3) lateral separation of the coiled chromonemata.

* Aided by a grant from the Rockefeller Foundation to Dr. C. W. Metz.

[1] v. Herwerden, M. A., *Anat. Anz.*, **36**, 193–207 (1910).
[2] Kaufmann, B. P., *Am. Nat.*, **65**, 555–558 (1931).
[3] Sinotô, Y., and Yuasa, A., *Jap. Jour. of Gen.*, **10**, 245–248 (1935).
[4] Koltzoff, N., *Science*, **80**, 312–313 (1934).
[5] Bridges, C. B., *Jour. of Hered.*, **26**, 60–64 (1935).
[6] Painter, T., *Am. Nat.*, **73**, 315–330 (1939).
[7] Bauer, H., *Zeit. Zellf.*, **23**, 280–313 (1935).
[8] Metz, C. W., *Cold Spring Harbor Symposia on Quantitative Biology*, **9**, 23–36 (1941).
[9] Alverdes, F., *Arch. Zellf.*, **9**, 168–204 (1912).

[10] Tänzer, E., *Zeit. wiss. Zool.*, **119**, 114–153 (1921).
[11] Dawydov, W., *Zeit. Zellf.*, **10**, 625–641 (1930).
[12] Kaufmann, B. P., *Am. Nat.*, **65**, 280–282 (1931).
[13] Koshy, T. K., *Jour. Roy. Microsc. Soc.*, **54**, 104–120 (1934).
[14] Naithani, S. P., *Ann. Bot. N. S.*, **1**, 257–276 (1937).
[15] Swanson, C. P., *Am. Jour. Bot.*, **30**, 422–428 (1943).
[16] Ris, H., (in press).
[17] Cooper, K. W., these PROCEEDINGS, **24**, 452–458 (1938).
[18] Ris, H., *Jour. Exp. Zool.*, **90**, 267–322 (1942).
[19] Doyle, W. L., and Metz, C. W., *Biol. Bull.*, **69**, 126–135 (1935).
[20] Mirsky, A. E., and Pollister, A. W., these PROCEEDINGS, **28**, 344–352 (1942).

THE SPREAD OF AN EPIDEMIC

BY EDWIN B. WILSON AND JANE WORCESTER

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated September 11, 1945

The Frost-Soper theory of the rise and fall of an epidemic in a population[1] can at best represent only a part of the phenomenon of epidemics, for they spread from population to population as well as from individual infectious cases to susceptible individuals within a population. The fundamental hypothesis or law, namely,

$$C(t) = rS(t)C(t - \tau) \tag{1}$$

where C is the new-case rate, S the number of susceptibles, r a factor of proportionality and τ the time lag between infection and infectiousness may be very naturally extended to represent what hypothetically might happen between two populations S and S' of susceptibles by writing

$$C(t) = rS(t)C(t - \tau) + r''S(T)C'(t - \tau), \tag{2}$$

$$C'(t) = r'S'(t)C'(t - \tau) + r'''S'(t)C(t - \tau). \tag{2'}$$

It would indeed be only natural to argue that the new-case rate C among the susceptibles S should be the sum of what it would be due to the infectious rate within S and of a similar term due to the infectious rate in the other population of which a number of infectious individuals proportional to the total number within S' come in contact with the susceptibles S.

It may first be observed that if the two susceptible populations S and S' were merely ideal subdivisions of one and the same population S_0 so that $S = pS_0$, $S' = qS_0$, $p + q = 1$, and $C = pC_0$, $C' = qC_0$, the rate of intermixture of the infectious $C(t - \tau)$ with the susceptibles S or S' being the same as the rate of intermixture of the infectious $C(t - \tau)$ with S_0, and the

same for $C'(t-\tau)$, we should have $r = r' = r'' = r'''$ and the two equations would become

$$pC_0(t) = rp^2S_0(t)\ C_0(t-\tau) + rpqS_0(t)\ C_0(t-\tau),$$
$$qC_0(t) = rq^2S_0(t)\ C_0(t-\tau) + rpqS_0(t)\ C_0(t-\tau).$$

Each of these equations is equivalent to

$$C_0(t) = rS_0(t)C_0(t-\tau)$$

which is that of the whole epidemic within the whole population of susceptibles.

If, however, one should at any instant actually separate the population S_0 of susceptibles into two parts $S = pS_0$, $S' = qS_0$ and the infectious cases $C_0(t-\tau) = pC_0(t-\tau)$, $C'(t-\tau) = qC_0(t-\tau)$, imagining that the rate of intermixture of each set of infectious cases with its corresponding set of susceptibles has not been thereby changed but that the intermixture of the infectious $C(t-\tau)$ with susceptibles $S'(t)$ and of $C'(t-\tau)$ with $S(t)$ has been completely stopped by this separation one would have $r = r'$ but $r'' = r''' = 0$ and the two equations for the two separate epidemics must relapse to the form

$$C(t) = rS(t)C(t-\tau) \qquad \text{and } C'(t) = rS'(t)C'(t-\tau).$$

As the values of $S(t)$ and $C(t-\tau)$ are initially pS_0 and $pC_0(t-\tau)$, the value of the new-case rate $C(t)$ has been multiplied by the factor p^2 instead of by p and $C'(t)$ has been multiplied by q^2 and the total new-case rate $C(t) + C'(t)$ has initially been multiplied by $p^2 + q^2$ which at most is equal to $1/2$. The subsequent course of the two epidemics, each being regulated by its own equation, will not result in the total subsequent number of cases being the fraction $p^2 + q^2$ of the value of the total number which would have occurred if the separation had not been made, but may be something very different.[2] The principle, however, that one tends to break up an epidemic by separating the susceptibles into smaller groups even if the intimacy of intermixture of the infectious and the susceptibles within each group is not reduced is well known.

It is obvious that any argument which would lead to equations (2) and (2′) for the regulation of an *a priori* epidemic in two intercommunicating populations would immediately generalize into the formula

$$C_i(t) = S_i(t) \sum_{j=1}^{n} r_{ij}C_j(t-\tau), \qquad i = 1, 2, \ldots, n, \qquad (3)$$

for any one of n intercommunicating populations. If there are no new susceptibles coming into the population $C = -dS/dt$ and the integral

$$\log S_i(t) = \sum_{j=1}^{n} r_{ij}S_j(t-\tau) - K_i, \qquad i = 1, 2, \ldots, n, \quad (4)$$

may be obtained at once and, given the value of r_{ij} and of $S_i(t_0)$ and $S_i(t_0 - \tau)$ at the start, one can compute step by step all the n epidemics in the populations. It would seem that to consider intercommunicating populations under such an *a priori* law would be one step toward realism beyond the assumption of the simple mass law for a single population.

We shall take as an illustration two populations $S = 2000$, $S' = 1200$, $r = r' = 0.001$, $r'' = r''' = 0.0001$, so that the rate of intermixture of infectious from one population and susceptibles of the other is only one-tenth the rate of intermixture of the infectious of either population with that population. We start with the next value of S as 1998 and of S' as unchanged at 1200 to determine the constants. If we use $x = 0.001S$ and $x' = 0.001S'$ we have

$$\log x_{T+1} = x_T + 0.1\, x'_T - 1.42785,$$

$$\log x'_{T+1} = x'_T + 0.1\, x_T - 1.21768.$$

In successive generations the cases in the first and second populations and in the total are given in the first three columns of table 1.

From the above calculation made with the law of mass action but with allowance for a smaller rate of contact between the two populations than within each population, we may consider three theoretical epidemics, either that in the first or that in the second or that in the total population, and this is, in a simple idealized form, just what is actually done in the discussion of epidemics of infectious disease where the health officer may be interested in part of a population as well as in the total. If we consider, as Soper did, that the ratio of case rates at the center of successive incubation periods may be approximated by the ratio of the cases during those successive periods, we may obtain from the values of S (not given in table 1) and from the cases as given[3] values for m on the assumption that $p = 1$ (in our previous notation); these values of m are given in the last three columns of table 1 for the epidemics defined by the respective first three columns.

What is observed from the table is that the three epidemics are quite typical but that if any one of the three be used to determine the value of m, assuming $p = 1$, the value of m is not constant; for the epidemic in the first population m decreases throughout (the infectivity measured as the reciprocal of m increases), for that in the second population m increases throughout (the infectivity decreases), and for that in the total m increases to a maximum and then decreases (the infectivity decreases to a minimum and then increases). If, relying on the simple equation $C(t)/C(t - \tau) = S(t)/m$ one should take as the value of m the number of susceptibles when $C(t) = C(t - \tau)$, estimating the case rates as best he could, he would find something like $m_{\mathrm{I}} = 985$, $m_{\mathrm{II}} = 830$, $m_{\mathrm{tot.}} = 1880$. If one should get the value of m/p from the ratio of total cases squared to eight times the peak

cases he would find something like these three respective values: 1050, 520, 1590. With the values of m found previously he would then estimate $p_{\mathrm{I}} = 0.94$, $p_{\mathrm{II}} = 1.6$, $p_{\mathrm{tot.}} = 1.2$. Thus he would infer that $p = 1$ did not satisfy the relations. If these values of p were used to compute m the results would be different from those in the table.[4]

TABLE 1

A THEORETICAL EPIDEMIC IN EACH OF TWO INTERCOMMUNICATING POPULATIONS AND IN THE WHOLE POPULATION WITH THE VALUES OF SOPER'S m FOR EACH

CASES I	CASES II	TOTAL	m_{I}	m_{II}	$m_{\mathrm{tot.}}$
2.1	0.0	2.0	...	...	...
4.0	0.2	4.2	1001	...	1523
8.0	0.8	8.7	993	...	1547
15.9	1.9	17.8	989	...	1566
31.5	4.1	35.6	988	543	1578
60.8	8.6	69.4	987	574	1593
112.3	17.3	129.6	986	593	1617
190.3	32.8	223.2	986	616	1660
277.2	57.3	334.6	986	648	1730
319.9	87.8	407.8	985	698	1824
274.0	111.7	385.7	974	765	1899
174.7	113.9	288.6	951	836	1900
89.9	94.0	183.9	918	893	1841
41.5	65.5	107.0	876	930	1776
18.7	40.6	59.3	835	950	1732
8.5	23.4	32.0	796	961	1705
4.0	13.0	17.0	787	969	1693
1.9	7.0	8.9	...	973	1686
1.0	3.7	4.7	...	973	1681
0.5	2.0	2.4	...	979	1678
0.2	1.0	1.3	...	...	..
0.1	0.6	0.7	...	...	...

From such a simple hypothetical example one may infer that even if epidemics in single homogeneous populations when isolated should be accountable by the law of mass action with $p = 1$, epidemics in groups of intercommunicating populations could not be so accounted for. As any real population is in fact not homogeneous but at best must be considered as formed of parts with different rates of intermixing of infectious and susceptibles both as within and as between parts, there is ample justification for generalizing the law of mass action to other values of p than 1 as an empirical law if thereby the observed relations are notably better reproduced.

We may observe that if populations are intercommunicating the whole question of periodicity becomes more complicated than in a single homo-

geneous population. If we take equations (2) and (2′) and replace C by $A - dS/dt$ where A is the rate of accession of susceptibles we have

$$A - \frac{dS}{dt} = S\left[r\left(A - \frac{dS}{dt}\Big|_{t-\tau}\right) + r''\left(A' - \frac{dS'}{dt}\Big|_{t-\tau}\right)\right],$$

$$A' - \frac{dS'}{dt} = S'\left[r'\left(A' - \frac{dS'}{dt}\Big|_{t-\tau}\right) + r'''\left(A - \frac{dS}{dt}\Big|_{t-\tau}\right)\right].$$

For the steady state to be possible in each population with $S = m$ and $S' = m'$ it is necessary that

$$A = m(rA + r''A') \qquad \text{and } A' = m'(r'A' + r'''A).$$

For an infinitesimal disturbance one has $S = m(1 + u)$, $S' = m'(1 + u')$. The symbolic form of the linear equations regulating the infinitesimal epidemics is therefore

$$[D - rmDe^{-\tau D} + rA + r''A']u - r''m'De^{-\tau D}u' = 0$$
$$-r'''mDe^{-\tau D}u + [D - r'm'De^{-\tau D} + r'A' + r'''A]u' = 0.$$

The periods (and dampings) will thus be had from the equation

$$[D - rmDe^{-\tau D} + rA + r''A'][D - r'm'De^{-\tau D} + r'A' + r'''A] - r''r'''mm'D^2e^{-2\tau D} = 0.$$

The simplest approximation to make would be to assume $e^{-\tau D} = 1 - \tau D$. This would lead to a biquadratic equation for D and thus normally to a pair of periods with different periodic times. Presumably the damping would be neglected. It is well known that the superposition of a pair of periodic terms may make a very irregular resultant variation.

Now, the variation of measles over the different wards of a city or over the different suburbs of a large metropolitan area is such, on the record, as to show both some degree of independence of the disease in the different sections and some degree of interdependence between sections.[5] While the general theory, whether of rise or fall or of periodicity, built upon some simple empirical law may throw much light on the observed phenomena, no such theory can be expected to fit them exactly. It seems to be indicated that one must study in detail the laws of spread, whether in the field or in the laboratory, before he can be finally satisfied with the theoretical discussion of epidemics.

[1] Wilson, E. B., and Burke, M. H., these PROCEEDINGS, **28**, 361–367 (1942); **29**, 43–48 (1943); and Wilson, E. B., and Worcester, J., *Ibid.*, **30**, 37–44 and 264–269 (1944); **31**, 24–34, 109–116, 142–147, 203–208, 294–298 (1945).

[2] If we take an epidemic starting with 2000 susceptibles and $r = 0.001$ so that the next number of susceptibles is 1998, it being assumed that there are no new susceptibles coming into the population, the rule for successive generations is $\log x_{T+1} = x_T - 1.30785$.

where $x_T = 0.001S_T$. The epidemic then goes on with these successive values of S, the cases being the differences between the values of S_0,

$S = 2000.00,$	$1998.00,$	$1994.01,$	$1986.08,$	$1970.40,$
Cases $2.00,$	$3.99,$	$7.93,$	$15.68.$	

If the calculation be carried on indefinitely we should have a final asymptotic $S = 406$ and total cases of 1594. If, however, the population be cut in two at the point reached and we consider $S_T = 985.20$ susceptibles with 7.84 infectious cases and $r = 0.001$, the number of new cases would be 7.72 with $S_{T+1} = 977.48$ so that $\log x_{T-1} = x_T - 1.00797$ would be the equation to use in the stepwise calculation. The argument could also be given in this form: The new number of cases if the separation had not been made would have been 30.66; with the separation it should have been cut in half to 15.33 and for each half it would have been 7.66; then taking $S_T = 985.20$, $S_{T+1} = 977.54$ we should determine $\log x_{T+1} = x_T - 1.00792$ for the stepwise calculation. Probably the best value of x_{T+1} would be between 0.97748 and 0.97754 and the best value of the constant between 1.00797 and 1.00792. To get the best new equation for each half of the epidemic, particularly in cases where the separation is made when the ratio of new cases to susceptibles is higher than in this example and where the two modes of calculation would differ more seriously, would require very careful analysis involving expansion into series in the time so as to distinguish adequately between cases and case rates. If the calculation of the example just taken be made with the second equation obtained the total number of cases in each half after the time of separation is 105 instead of 1565 for the whole after the corresponding time; it will be noted that $2 \times 105 = 210$ is decidedly less than $p^2 + q^2 = {}^1/_2$ of 1565. In fact the separation has stopped the epidemic so completely that the number of cases decreases from the time of separation.

[3] There are a variety of formulae which may be developed for getting the ratio of two ordinates in terms of successive areas. The one used by Soper is perhaps the simplest. As, however, the epidemic case-rate curve for a single population much resembles the sech^2 curve we might consider that for the ratio $C(t - \tau/2)/C(t + \tau/2)$ one could fit a sech^2 curve to C from $t - 3\tau/2$ to $t + 3\tau/2$ using K_-, K_0 and K_+ as the cases in the intervals $t - 3\tau/2$ to $t - \tau/2$, $t - \tau/2$ to $t + \tau/2$, and $t + \tau/2$ to $t + 3\tau/2$. The simple resulting formula is

$$\frac{C(t - \tau/2)}{C(t + \tau/2)} = \frac{m(t + \tau/2)}{S(t + \tau/2} = \frac{K_-}{K_+}\frac{K_0 + K_+}{K_0 + K_-}$$

and it has been used for the actual calculation of m instead of Soper's. The difference between the two is, however, tolerably small.

[4] We might, as in a previous article, keep p constant and compute the variation of m throughout the epidemic, or we might keep m constant and compute the variation of p, or finally we might let both m and p be determined from successive pairs of data. In using data from actual epidemics we found a great deal of scattering in the successive values of m and p but apparently a tendency for m to decrease throughout the epidemic and for p first to decrease from large values and then to increase to large values. If we make the corresponding calculations on these theoretical epidemics we find that m remains practically constant throughout the major course of the epidemic (at the ends enough places were lost to make the figures unreliable) and the value of p decreases throughout the epidemic whether this be that in either part of the population or that in the whole. We have assumed there are no recruits to the susceptibles, for without this assumption we cannot get exact integrals; clearly when there are recruits, as there probably are in real epidemics, they may well make a considerable difference with calcu-

lations when the number of new cases is less or only slightly greater than the number of new susceptibles during each incubation period.

[8] Wilson, E. B., Bennett, C., Allen, M., and Worcester, J., *Proc. Amer. Phil. Soc.*, **80**, 357–476 (1939). See particularly Appendix II, Localness of Measles, 469–476.

A TOPOLOGY FOR THE SET OF PRIMITIVE IDEALS IN AN ARBITRARY RING

BY N. JACOBSON

DEPARTMENT OF MATHEMATICS, JOHNS HOPKINS UNIVERSITY

Communicated July 11, 1945

1. In a recent paper we have called a ring $\mathfrak{A}$ primitive if $\mathfrak{A}$ contains a maximal right ideal $\mathfrak{J}$ whose quotient $\mathfrak{J}:\mathfrak{A} = 0$.[1] In general if $\mathfrak{J}$ is any right ideal $\mathfrak{J}:\mathfrak{A}$ is the totality of elements b such that $xb \in \mathfrak{J}$ for all x in $\mathfrak{A}$. $\mathfrak{J}:\mathfrak{A}$ is a two-sided deal and if $\mathfrak{A}$ has an identity, $\mathfrak{J}:\mathfrak{A}$ is the largest two-sided ideal of $\mathfrak{A}$ contained in $\mathfrak{J}$. The primitive rings appear to play the same rôle in the general structure theory of rings that is played by simple rings in the classical theory of rings that satisfy the descending chain condition for one-sided ideals. Corresponding to the Wedderburn-Artin structure theorem on simple rings satisfying the descending chain condition we have the theorem that if $\mathfrak{A}$ is a primitive ring $\neq 0$, $\mathfrak{A}$ is isomorphic to a dense ring of linear transformations in a suitable vector space over a division ring.

We shall call a two-sided ideal $\mathfrak{B}$ in $\mathfrak{A}$ a *primitive ideal* if $\mathfrak{B} \neq \mathfrak{A}$ and $\mathfrak{A} - \mathfrak{B}$ is a primitive ring. It is known that if $\mathfrak{A}$ is not a radical ring then $\mathfrak{A}$ contains primitive ideals. Moreover, in this case the intersection $\Pi\mathfrak{B}$ of all the primitive ideals in $\mathfrak{A}$ coincides with the radical $\mathfrak{R}$ of $\mathfrak{A}$. In particular if $\mathfrak{A}$ is semi-simple, $\Pi\mathfrak{B} = 0$. If $\mathfrak{A}$ is a ring with an identity, $\mathfrak{A}$ is not a radical ring. Hence if $\mathfrak{C}$ is any two-sided ideal $\neq \mathfrak{A}$ in $\mathfrak{A}$, $\mathfrak{A} - \mathfrak{C}$ contains a primitive ideal $\mathfrak{B} - \mathfrak{C}$. Since $\mathfrak{A} - \mathfrak{B} \cong (\mathfrak{A} - \mathfrak{C}) - (\mathfrak{B} - \mathfrak{C})$, $\mathfrak{B}$ is primitive in $\mathfrak{A}$. Thus if $\mathfrak{A}$ is a ring with an identity any two-sided ideal $\mathfrak{C} \neq \mathfrak{A}$ can be imbedded in a primitive ideal.

Any commutative primitive ring is a field. Consequently any primitive ideal in a commutative ring is maximal. On the other hand, in a non-commutative ring there may exist primitive ideals that are not maximal. For example, the ring $\mathfrak{L}$ of all linear transformations in an infinite dimensional vector space $\mathfrak{R}$ over a division ring is primitive. Hence (0) is a primitive ideal. However, $\mathfrak{L}$ contains as a proper two-sided ideal the set $\mathfrak{F}$ of finite valued linear transformations in $\mathfrak{R}$.

A simple ring is either primitive or a radical ring. In particular any simple ring with an identity is primitive. These remarks imply that a maximal two-sided ideal $\mathfrak{B}$ such that $\mathfrak{A} - \mathfrak{B}$ is not a radical ring is primitive. If $\mathfrak{A}$ is a ring with an identity, any maximal two-sided ideal in $\mathfrak{A}$ is primitive.

In this note we shall define a topology for the set of primitive ideals of any ring. The space determined in this way appears to be an important invariant of the ring. We hope to discuss its rôle in the general structure theory in greater detail at a later date.

2. Let S be the set of primitive ideals in the ring $\mathfrak{A}$. S is vacuous if and only if $\mathfrak{A}$ is a radical ring. If A is a non-vacuous subset of S we let $\mathfrak{D}_A$ denote the intersection of all the primitive ideals $\mathfrak{B} \in A$. We now define the closure $\overline{A}$ of A to be the totality of primitive ideals $\mathfrak{C}$ such that $\mathfrak{C} \geqq \mathfrak{D}_A$. It is clear that

1. $\bar{A} \geqq A$.
2. $\overline{\bar{A}} = \bar{A}$.

We wish to prove next that

3. $\overline{A \vee B} = \bar{A} \vee \bar{B}$.[2]

Proof. Let $\mathfrak{B} \in \bar{A} \vee \bar{B}$, say $\mathfrak{B} \in \bar{A}$. Then $\mathfrak{B} \geqq \mathfrak{D}_A$. Hence $\mathfrak{B} \geqq \mathfrak{D}_A \wedge \mathfrak{D}_B = \mathfrak{D}_C$ where $C = A \vee B$. Thus $\mathfrak{B} \in \overline{A \vee B}$. Suppose next that $\mathfrak{B} \in \overline{A \vee B}$. Then $\mathfrak{B} \not\geqq \mathfrak{D}_A$ and $\mathfrak{B} \not\geqq \mathfrak{D}_B$. We consider now the primitive ring $\mathfrak{A} - \mathfrak{B}$. The two-sided ideals $(\mathfrak{D}_A + \mathfrak{B}) - \mathfrak{B}$ and $(\mathfrak{D}_B + \mathfrak{B}) - \mathfrak{B}$ are $\neq 0$ in $\mathfrak{A} - \mathfrak{B}$. Hence the product $[(\mathfrak{D}_A + \mathfrak{B}) - \mathfrak{B}][(\mathfrak{D}_B + \mathfrak{B}) - \mathfrak{B}] \neq 0$.[3] It follows that $\mathfrak{D}_A \mathfrak{D}_B \not\leqq \mathfrak{B}$. Hence also $\mathfrak{D}_A \wedge \mathfrak{D}_B \not\leqq \mathfrak{B}$ and so $\mathfrak{B} \in \overline{A \vee B}$.

For the vacuous set ω we define

4. $\bar{\omega} = \omega$.

The properties 1–4 show that *S is a topological space relative to the closure operation $A \to \bar{A}$.* In the special case of a Boolean ring this topology is due to Stone.[4] It has also been introduced by Gelfand and Silov in commutative normed rings.[5] We shall call the topological space S the *structure space* of the ring $\mathfrak{A}$.

If $\mathfrak{B}$ is a primitive ideal in $\mathfrak{A}$, $\mathfrak{B}$ is a point in S. The closure of the set $\{\mathfrak{B}\}$ is the totality of primitive ideals $\mathfrak{C}$ such that $\mathfrak{C} \geqq \mathfrak{B}$. Hence if $\{\overline{\mathfrak{B}_1}\} = \{\overline{\mathfrak{B}_2}\}$ then $\mathfrak{B}_1 = \mathfrak{B}_2$. This shows that S is a T_0-space. In general S is not a T_1-space. For we have seen that there exist primitive rings $\mathfrak{A}$ that are not simple. If $\mathfrak{A}$ is a ring with an identity of this type and $\mathfrak{B}$ is a proper two-sided ideal in $\mathfrak{A}$, $\mathfrak{B}$ can be imbedded in a primitive ideal $\mathfrak{C}$. Then $\{\overline{(0)}\}$ contains $\mathfrak{C} \neq 0$. The subspace M of S of primitive ideals that are maximal is clearly a T_1-space. If $\mathfrak{A}$ is commutative $S = M$ is a T_1-space. However, even in this case S need not be a T_2- (or Hausdorff) space. An example of a normed ring of this type has been given by Gelfand and Silov.[5] A simpler one is the following:

Example. Let $\mathfrak{A} = J$ the ring of integers. The primitive ideals are the prime ideals (p). Since the intersection of an infinite number of prime ideals is the 0-ideal, $\bar{A} = S$ for any infinite set A. If A is finite, $\bar{A} = A$. Hence the open sets $\neq \omega, \neq S$ are the complements of finite sets. Any two open sets $\neq \omega$ have a non-vacuous intersection and so the Hausdorff separation property does not hold.

3. Let $\mathfrak{A}_1$ be an arbitrary two-sided ideal in $\mathfrak{A}$ and let S_1 denote the closed set in S consisting of the primitive ideals $\mathfrak{B}$ of $\mathfrak{A}$ that contain $\mathfrak{A}_1$. If $\mathfrak{B} \in S_1$ then $\mathfrak{B} - \mathfrak{A}_1$ is a primitive ideal in $\mathfrak{A} - \mathfrak{A}_1$ and any primitive ideal in $\mathfrak{A} - \mathfrak{A}_1$ is obtained in this way. The correspondence $\mathfrak{B} \to \mathfrak{B} - \mathfrak{A}_1$ is $(1-1)$ between the subspace S_1 of S and the structure space T of the ring $\mathfrak{A} - \mathfrak{A}_1$. Since this correspondence preserves intersection it is a homeomorphism between S_1 and T.

Suppose in particular that $\mathfrak{A}_1 = \mathfrak{N}$ the radical of $\mathfrak{A}$. Then every primitive ideal of $\mathfrak{A}$ contains $\mathfrak{N}$. Hence $S_1 = S$, and we see that *the structure space of $\mathfrak{A}$ is homeomorphic to the structure space of the semi-simple ring* $\mathfrak{A} - \mathfrak{N}$.

We return to the general case in which $\mathfrak{A}_1$ is arbitrary and we now consider the open set S_1' of primitive ideals that do not contain $\mathfrak{A}_1$. Let $\mathfrak{C} \in S_1'$. Then $\mathfrak{C} \wedge \mathfrak{A}_1$ is a two-sided ideal $\neq \mathfrak{A}_1$ in $\mathfrak{A}_1$ and $\mathfrak{A}_1 - (\mathfrak{C} \wedge \mathfrak{A}_1) \cong (\mathfrak{C} + \mathfrak{A}_1) - \mathfrak{C}$. The latter ring is a two-sided ideal in the primitive ring $\mathfrak{A} - \mathfrak{C}$. Hence it is primitive.[6] Thus $\mathfrak{C} \wedge \mathfrak{A}_1$ is in the structure space U of the ring $\mathfrak{A}_1$. Let A be a subset of S_1' and let $\mathfrak{C} \in \bar{A} \wedge S_1'$ so that $\mathfrak{C}$ is in the closure of A in the subspace S_1'. If $\mathfrak{D}_A$ is the intersection of the primitive ideals in A then $\mathfrak{C} \supseteq \mathfrak{D}_A$ but $\mathfrak{C} \not\supseteq \mathfrak{A}_1$. Let B be the subset of U of ideals $\mathfrak{B} \wedge \mathfrak{A}_1$ where $\mathfrak{B} \in A$. The intersection of all of these ideals is the ideal $\mathfrak{D}_A \wedge \mathfrak{A}_1$. Since $(\mathfrak{C} \wedge \mathfrak{A}_1) \supseteq (\mathfrak{D}_A \wedge \mathfrak{A}_1)$, $\mathfrak{C} \wedge \mathfrak{A}_1$ is in $\bar{B}$. Hence the mapping that we have defined between S_1' and the subspace of U is a continuous one.

4. Let $\{F_\alpha\}$ be a set of closed sets in the space S. As before let $\mathfrak{D}_{F_\alpha}$ denote the two-sided ideal of elements common to all the $\mathfrak{B} \in F_\alpha$. Suppose that the intersection $\Pi F_\alpha \neq \omega$ and let $\mathfrak{B}$ be a point in this intersection. Then $\mathfrak{B} \supseteq \mathfrak{D}_{F_\alpha}$ for all α. Hence $\mathfrak{B}$ contains the two-sided ideal $\Sigma\mathfrak{D}_{F_\alpha}$ generated by the $\mathfrak{D}_{F_\alpha}$. The converse follows by retracing the steps of this argument. We therefore have the

LEMMA 1. *If $\{F_\alpha\}$ is a set of closed sets in S, $\Pi F_\alpha \neq \omega$ if and only if $\Sigma\mathfrak{D}_{F_\alpha}$ can be imbedded in a primitive ideal.*

If $\mathfrak{A}$ is a ring with an identity any two-sided ideal $\neq \mathfrak{A}$ can be imbedded in a primitive ideal. Hence we have the

Corollary. If $\mathfrak{A}$ is a ring with an identity and $\{F_\alpha\}$ is a set of closed sets in S then $\Pi F_\alpha \neq \omega$ if and only if $\Sigma\mathfrak{D}_{F_\alpha} \neq \mathfrak{A}$.

If $\mathfrak{A}$ has an identity and $\Sigma\mathfrak{D}_{F_\alpha} = \mathfrak{A}$, $1 = d_1 + \ldots + d_r$ where $d_i \in \mathfrak{D}_{F_i}$. Hence also $\Sigma\mathfrak{D}_{F_i} = \mathfrak{A}$. The corollary therefore shows that if $\Pi F_\alpha = \omega$

then there is a finite set of closed sets F_i such that $\Pi F_i = \omega$. We therefore have the following

THEOREM 1. *The structure space of a ring with an identity is bicompact.*

Suppose now that $\mathfrak{A}$ is a semi-simple ring. Let $\mathfrak{A}$ be decomposable as a direct sum $\mathfrak{A}_1 \oplus \mathfrak{A}_2$ of the two-sided ideals $\mathfrak{A}_i \neq 0$. Let S_i be the closed subset of S consisting of the primitive ideals $\mathfrak{B}$ containing the ideal $\mathfrak{A}_i$. Since $\mathfrak{A} - \mathfrak{A}_1 \cong \mathfrak{A}_2$ a semi-simple ring, $S_1 \neq \omega$. Similarly $S_2 \neq \omega$. Also the semi-simplicity of $\mathfrak{A} - \mathfrak{A}_i$ implies that $\mathfrak{A}_i$ is the intersection of all the $\mathfrak{B}$ in S_i. Since $\mathfrak{A}_1 + \mathfrak{A}_2 = \mathfrak{A}$ the lemma implies that $S_1 \wedge S_2 = \omega$. Now let $\mathfrak{B}$ be any primitive ideal in $\mathfrak{A}$. We assert that either $\mathfrak{B} \geqq \mathfrak{A}_1$ or $\mathfrak{B} \geqq \mathfrak{A}_2$. For otherwise $\mathfrak{A}_i + \mathfrak{B} > \mathfrak{B}$. The ideals $(\mathfrak{A}_i + \mathfrak{B}) - \mathfrak{B}$ are $\neq 0$ in the primitive ring $\mathfrak{A} - \mathfrak{B}$. Since $[(\mathfrak{A}_1 + \mathfrak{B}) - \mathfrak{B}][(\mathfrak{A}_2 + \mathfrak{B}) - \mathfrak{B}] = 0$ this is impossible. We have therefore proved our assertion. Evidently it is equivalent to the relation $S_1 \vee S_2 = S$. Thus S is disconnected into the two components S_1 and S_2.

Conversely suppose that $S = S_1 \vee S_2$ where the S_i are closed sets $\neq \omega$ such that $S_1 \wedge S_2 = \omega$. We assume also that $\mathfrak{A}$ has an identity. Let $\mathfrak{A}_i = \mathfrak{D}_{S_i}$. Then $\mathfrak{A}_1 + \mathfrak{A}_2 = \mathfrak{A}$. Since $\mathfrak{A}$ is semi-simple $\mathfrak{A}_1 \wedge \mathfrak{A}_2 = \mathfrak{D}_S = 0$. If $\mathfrak{A}_i = 0$, $S = S_i = S_i$ contrary to $S_1 \neq \omega$ and $S_2 \neq \omega$. A part of our result is the following

THEOREM 2. *If $\mathfrak{A}$ is a semi-simple ring with an identity, $\mathfrak{A} = \mathfrak{A}_1 \oplus \mathfrak{A}_2$ where the $\mathfrak{A}_i$ are two-sided ideals $\neq 0$ if and only if S is not connected.*

5. If S' is a completely regular bicompact space and $\mathfrak{S}$ is the ring of real-valued (complex valued) continuous functions on S' then it has been shown by Gelfand and Silov that the structure space S of $\mathfrak{S}$ is homeomorphic to S'.[7] If S' is a bicompact totally disconnected space and $\mathfrak{S}$ is the ring of continuous functions on S' having values in the field of residues mod 2, then by a result of Stone's the structure space of $\mathfrak{S}$ is homeomorphic to S'.[8] We conclude this note by giving another example of this type based on an arbitrary totally disconnected bicompact space S' and an arbitrary division ring $\mathfrak{R}'$.

We consider any decomposition of S' into a finite number of components (non-overlapping open and closed sets) S'_i and we choose corresponding elements $k_i \in \mathfrak{R}'$. We define a function $f(x)$ by setting $f(x_i) = k_i$ for x_i in S'_i. A function of this type will be called a *finite decomposition function*. The totality $\mathfrak{S}$ of these functions is a ring under the ordinary operations of addition and multiplication. $\mathfrak{S}$ contains the subring $\mathfrak{K}$ of constant functions, isomorphic to $\mathfrak{R}'$ and $\mathfrak{S}$ is commutative if and only if $\mathfrak{R}'$ is commutative. The constant 1 acts as an identity in $\mathfrak{S}$. Since S' is totally disconnected, for any two points $a \neq b$ in S' there is an $f(x) \in \mathfrak{S}$ such that $f(a) \neq f(b)$. Let $\mathfrak{A}$ be any subring of $\mathfrak{S}$ having this property and containing $\mathfrak{K}$. We shall sketch a proof of the fact that the space M of maximal two-sided ideals of $\mathfrak{A}$ is homeomorphic to S'.

LEMMA 2. *If F' is a closed subset of S' and a is a point $\bar{\epsilon}$ F' then there exists a function $\varphi(x) \epsilon \mathfrak{A}$ such that $\varphi(y) = 0$ for all $y \epsilon F'$ but $(\varphi a) \neq 0$.*

Our assumptions imply that for each $y \epsilon F'$ there is an $f_y \epsilon \mathfrak{A}$ such that $f_y(y) = 0$ but $f_y(a) \neq 0$. The set $Z'(f_y)$ of zeros of f_y is open and the totality of these sets covers F'. Let $Z'(f_{y_1}), \ldots, Z'(f_{y_m})$ be a finite subset of these sets covering F'. Then $\varphi(x) = f_{y_1}(x) \ldots f_{y_m}(x)$ has the required properties.

If $a \epsilon S'$ we let $\mathfrak{B}_a$ denote the totality of functions $g(x) \epsilon \mathfrak{A}$ such that $g(a) = 0$. It is easy to see that $\mathfrak{B}_a$ is a maximal two-sided ideal in $\mathfrak{A}$ and that $\mathfrak{A} - \mathfrak{B}_a \cong \mathfrak{K}$. Our conditions imply that if $a \neq b$ then $\mathfrak{B}_a \neq \mathfrak{B}_b$.

LEMMA 3. *If $\mathfrak{B}$ is a two-sided ideal $\neq \mathfrak{A}$, then there exists a point a such that $g(a) = 0$ for all $g \epsilon \mathfrak{B}$.*

Let $Z'(f)$ be the set of zeros of f. $Z'(f)$ is closed. If the lemma is false, the intersection $\Pi Z'(f)$ for all $f \epsilon \mathfrak{B}$ is vacuous. Hence there is a finite, number of functions $f_1, \ldots, f_r$ in $\mathfrak{B}$ such that $\Pi Z'(f_i) = \omega$. Hence if a is any point of S' at least one of the functions f_i does not vanish at a. Let $k_{i1}, \ldots, k_{ni}$ be the non-zero values taken on by f_i and form the function $\psi_i(x) = (f_i(x) - k_{1i})(f_i(x) - k_{2i}) \ldots (f_i(x) - k_{ni})$ and $\psi(x) = \psi_1(x)\psi_2(x) \ldots \psi_r(x)$. Then $\psi(x) \equiv 0$. The form of $\psi(x)$ shows that $0 = \psi(x) = k + g(x)$ where $k \neq 0$ and $g(x) \epsilon \mathfrak{B}$. Hence $\mathfrak{B}$ contains a constant function $\neq 0$ and $\mathfrak{B} = \mathfrak{A}$ contrary to asumption.

This lemma shows that $\mathfrak{B} \leqq \mathfrak{B}_a$ for some a. If $\mathfrak{B}$ is maximal $\mathfrak{B} = \mathfrak{B}_a$. Thus the correspondence $a \rightarrow \mathfrak{B}_a$ is $(1 - 1)$ between S' and the space M of maximal two-sided ideals in $\mathfrak{A}$.

Let F' be a closed subset of S' and let F be the corresponding set in M. The intersection $\mathfrak{D}_F = \Pi \mathfrak{B}_a$ for all $a \epsilon F'$ is the set of functions f such that $f(a) = 0$ for all $a \epsilon F$. Let $\mathfrak{B}_b$ be a maximal two-sided ideal containing $\mathfrak{D}_F$. If $b \bar{\epsilon} F'$ there is a function ϕ such that $\phi(a) = 0$ for all a in F' but $\phi(b) \neq 0$. Then $\phi \epsilon \mathfrak{D}_F$ but $\bar{\epsilon} \mathfrak{B}_b$ contrary to $\mathfrak{B}_b \geqq \mathfrak{D}_F$. Hence $b \epsilon F'$ and $\mathfrak{B}_b \epsilon F$. Thus F is closed. Conversely let F be any closed set in M and let F' be the corresponding set in S'. It is easy to see that F' is the set of points y such that $f(y) = 0$ for all f in $\mathfrak{D}_F$. Thus F' is the intersection of closed sets and is therefore closed. Hence the correspondence $a \rightarrow \mathfrak{B}_a$ is a homeomorphism.

THEOREM 3. *Let S' be a totally disconnected bicompact space, $\mathfrak{K}'$ a division ring and $\mathfrak{A}$ a ring of $\mathfrak{K}'$-valued decomposition functions such that (1) $\mathfrak{A}$ contains the constants and (2) if $a \neq b$ in S', then $\mathfrak{A}$ contains a function f such that $f(a) \neq f(b)$. Then the space M of maximal two-sided ideals of $\mathfrak{A}$ is homeomorphic to S'.*

If $\mathfrak{K}'$ is commutative, $M = S$. It is an open question whether or not this is true in general. However, it is clear that $M = S$ since the intersection of the set of maximal two-sided ideals is (0).

Theorem 3 shows that two rings $\mathfrak{A}_1$ and $\mathfrak{A}_2$ of the type considered are not isomorphic unless the spaces S_1' and S_2' are homeomorphic. Since the division ring $\mathfrak{K}_i'$ is isomorphic to $\mathfrak{A}_i - \mathfrak{B}_i$ for any maximal two-sided ideal

$\mathfrak{B}_i$ in $\mathfrak{A}_i$, the isomorphism of $\mathfrak{A}_1$ and $\mathfrak{A}_2$ implies that of $\mathfrak{K}_1'$ and $\mathfrak{K}_2'$. If $\mathfrak{A}_i = \mathfrak{S}_i$ the complete ring of finite decomposition functions then the converse holds. Hence necessary and sufficient conditions that $\mathfrak{S}_1$ and $\mathfrak{S}_2$ be isomorphic are that S_1' and S_2' be homeomorphic and that $\mathfrak{K}_1'$ and $\mathfrak{K}_2'$ be isomorphic.

[1] "The Radical and Semi-simplicity for Arbitrary Rings," *Amer. Jour. Math.*, **67**, 300–320 (1945). The results that we state without proof in this section can be found in the above-mentioned paper.

[2] We use the notations $\vee$ and $\wedge$, respectively, for the logical sum and logical product of a finite number of sets.

[3] Loc. cit. in reference 1, p. 316.

[4] "Applications of the Theory of Boolean Rings to General Topology," *Trans. Amer. Math. Soc.*, **41**, 375–481 (1937).

[5] "Über verschiedene Methoden der Einführing der Topologie in die Menge der maximalen ideale eines normierten Ringes," *Math. Sbornik*, **51**, 25–39 (1941).

[6] Loc. cit. in reference 1, p. 313.

[7] Loc. cit. in reference 5, pp. 27–30.

[8] Loc. cit. in reference 4, p. 380. It is assumed here that the field of residues mod 2 is endowed with a T_1-topology. This means that each point is both open and closed.

CONVERSE OF PTOLEMY'S THEOREM ON STEREOGRAPHIC PROJECTION

By Edward Kasner and John De Cicco

Departments of Mathematics, Columbia University and Illinois Institute of Technology

Communicated August 22, 1945

1. The Problem.—The famous theorem of Ptolemy on stereographic projection states that the perspectivity of a sphere from a point of the sphere upon a plane perpendicular to the diameter of the sphere determined by the given point is conformal. The question naturally arises as to the existence of any other surfaces for which there exists a perspectivity upon a plane which is conformal. We prove that there does not exist any other surface with this property (except for the obvious case of a parallel plane). Our result may be stated as follows:

THEOREM. *If a perspectivity of a surface from a fixed point to a fixed plane is conformal, then the surface must be a sphere; furthermore, the sphere must pass through the fixed point and have its center on the perpendicular from the fixed point to the fixed plane.*

Thus the only perspective conformalities upon a plane are Ptolemy's stereographic projection of the sphere (or the limiting case of the plane).

2. Beginning of the Proof of the Theorem.—Let (x, y, z) denote cartesian

coördinates of a point and let $p = \partial z/\partial x$ and $q = \partial z/\partial y$. Then $(p, q, -1)$ are the direction numbers of any normal to a surface.

We shall take the origin as the center of perspectivity and $z = c \neq 0$ as the given plane. Thus (X, Y, c) are the cartesian coördinates of any point in the given plane. Let the equation of the unknown surface be $z = f(x, y)$.

The perspectivity from the given point $(0, 0, 0)$ of the surface $(x, y, z = f(x, y))$ upon the given plane (X, Y, c) is given by

$$X = \frac{cx}{z}, \qquad Y = \frac{cy}{z}. \tag{1}$$

The differentials of this point transformation are

$$dX = \frac{c}{z^2}[(z - xp)dx - xqdy], \qquad dY = \frac{c}{z^2}[-ypdx + (z - yq)dy]. \tag{2}$$

The jacobian J of our transformation is

$$J = \frac{c^2}{z^4}[(z - xp)(z - yq) - xypq] = \frac{c^2}{z^3}(z - xp - yq). \tag{3}$$

The vanishing of the jacobian shows that $z = f(x, y)$ is a homogeneous function of first degree, according to Euler's theorem on homogeneous functions. This means that the surface cannot be a cone with vertex at the origin. Thus since we require regions to correspond, *the plane $z = 0$ and any cone with vertex at origin must be omitted from consideration.*

By equation (2), it is found that the linear element dS in the plane $z = c$, is given by

$$dS^2 = dX^2 + dY^2 = \frac{c^2}{z^4}\begin{bmatrix} \{z^2 - 2xzp + (x^2 + y^2)p^2\}dx^2 + \\ 2\{-z(yp + xq) + (x^2 + y^2)pq\}\,dxdy \\ + \{z^2 - 2yzq + (x^2 + y^2)q^2\}dy^2 \end{bmatrix}. \tag{4}$$

On the other hand, the linear element of the surface $z = f(x, y)$ is given by

$$ds^2 = (1 + p^2)dx^2 + 2pqdxdy + (1 + q^2)dy^2. \tag{5}$$

Since we require that the perspectivity of the surface $z = f(x, y)$ upon the plane $z = c$, be conformal, the scale $\sigma = dS/ds$ is independent of direction. Hence from equations (4) and (5), we must have

$$\frac{z^2 - 2xzp + (x^2 + y^2)p^2}{1 + p^2} = \frac{-z(yp + xq) + (x^2 + y^2)pq}{pq} = \frac{z^2 - 2yzq + (x^2 + y^2)q^2}{1 + q^2}. \tag{6}$$

Subtracting z^2 from each ratio of this proportion, we find

$$\frac{p[-2xz + (x^2 + y^2 - z^2)p]}{1 + p^2} = \frac{-z(yp + xq) + (x^2 + y^2 - z^2)pq}{pq} = \frac{q[-2yz + (x^2 + y^2 - z^2)q]}{1 + q^2}. \quad (7)$$

Let M and N denote the quantities (linear in p and q)

$$M = -2xz + (x^2 + y^2 - z^2)p, \quad N = -2yz + (x^2 + y^2 - z^2)q. \quad (8)$$

The equation (7) may be written as

$$\frac{pM}{1 + p^2} = \frac{qM + pN}{2pq} = \frac{qN}{1 + q^2}. \quad (9)$$

3. The Special Cases, Real and Imaginary.—If p or q is zero, then both are zero, and we have a parallel plane as a solution. For if $p = 0$, then $qM = 0$. If $q \neq 0$, then $M = -2xz = 0$ which is impossible. Hence from $p = 0$, follows the fact that $q = 0$ which proves our statement. The argument when $q = 0$ is similar. Henceforth we may assume that neither p nor q is zero.

If the quantity $p = \pm i$, then the quantity $q = \pm 1$ (or if $p = \pm 1$, then $q = \pm i$) and the solution of equation (9) is an imaginary plane through the origin which is excluded from consideration since the perspectivity (1) is degenerate in this case. Let us suppose that $p = i$. From equation (9), we find $-2qM = 0$. Since $q \neq 0$, it follows that $M = -2xz + i(x^2 + y^2 - z^2) = 0$. This yields the imaginary planes $z = ix \pm y$, which are solutions of equation (9). However, this is excluded since the perspectivity (1) is degenerate in this case. A similar argument will prove our statement for $p = -i$ or $q = \pm i$.

Let us suppose $p = 1$. From equation (9), we find immediately that $N = 0$ and hence either $M = -2xz + x^2 + y^2 - z^2 = 0$ or $1 + q^2 = 0$. Now the condition $M = -2xz + x^2 + y^2 - z^2 = 0$ yields a contradiction; for the partial derivative q from this is $q = y/(x + z)$ and from $N = 0$, we find it to be $q = y/x$. The equality of these expressions yields the contradiction $yz = 0$. Hence when $p = 1$, we must have $1 + q^2 = 0$. This gives the imaginary planes $z = x \pm iy$, which are omitted since the perspectivity (1) is degenerate. The argument is similar when $p = -1$ or $q = \pm 1$. This completes the proof of the preceding italicized statement. Henceforth we may assume that neither p nor q assumes any of the five values $0, \pm i, \pm 1$.

We shall consider as our final imaginary case, the condition $p^2 + q^2 = 0$. This leads to the imaginary planes $z = a(x \pm iy)$ through the origin, which are omitted from consideration since these make the perspectivity (1) degenerate. A solution of this partial equation of first order is $z = f(u)$ where $u = x + iy$

(the case where $z = f(v)$ where $v = x - iy$, is similar). Hence $z = f(u)$, $p = f_u$, $q = ip = if_u$. The equations (9) then become

$$\frac{M}{1+p^2} = \frac{iM+N}{2ip^2} = \frac{iN}{1-p^2} \quad . \tag{10}$$

because $p \neq 0$. Since $p \neq \pm 1, \pm i$, these equations are equivalent to the single condition

$$i(p^2-1)M - (p^2+1)N = 0. \tag{11}$$

Substituting equation (8) into this and simplifying, we obtain

$$(uv - z^2)p + z(up^2 - v) = 0. \tag{12}$$

Since z and p are both functions of u only, this identity in v leads to the single condition $uf_u(u) - f(u) = 0$. This yields the solution $z = au = a(x + iy)$, which is an imaginary plane through the origin. As stated above, this must be omitted from consideration as it makes the perspectivity (1) degenerate. Henceforth p and q cannot be any of the five numbers 0, $\pm i$, ± 1, and they must not satisfy the relation $p^2 + q^2 = 0$.

4. *The General Case.*—The equations (9) are now equivalent to

$$q(p^2-1)M = p(p^2+1)N, \qquad p(q^2+1)M = q(p^2+1)N. \tag{13}$$

The determinant of these equations in M and N is $-(p^2+1)(p^2+q^2)$. Since this cannot be zero (as we have already treated these various cases), it follows that both M and N are zero. Hence from equation (8), we find

$$p = \frac{2xz}{x^2+y^2-z^2}, \qquad q = \frac{2yz}{x^2+y^2-z^2}. \tag{14}$$

The Pfaffian, $dz = pdx + qdy$, may be written in the form

$$2xzdx + 2yzdy - (x^2+y^2-z^2)dz = 0. \tag{15}$$

This can be made exact by dividing by $z(x^2+y^2+z^2)$. The complete integral is

$$x^2 + y^2 + z^2 = \text{const. } z. \tag{16}$$

These are spheres passing through the origin and with axis as the z-axis.

This completes the proof of our converse of the theorem of Ptolemy on stereographic projection. We have shown that the sphere is the only surface admitting a perspective conformality.

Elsewhere we shall discuss the possibility of two surfaces (neither one plane) admitting conformal perspectivity, a much more difficult problem.

5. *Other Characterizations by Scale Curves.*—The authors have given other characteristic properties of this stereographic map by means of scale curves. These are defined by the $\sigma = dS/ds = \text{const.}$ In any conformal

map upon a plane, there are ∞^1 scale curves. It develops that the most famous maps, namely, Mercator projection, stereographic (Ptolemy) and Lambert projections, have certain common characteristic properties.

The only conformal maps of a sphere upon a plane with isothermal scales are the Mercator projections (with straight scales), and the Ptolemy stereographic and Lambert projections (with circular scales), followed by a similitude in the given plane.

If ∞^1 geodesics of the sphere are mapped conformally upon a plane such that they are represented by straight lines, then the maps are stereographic (Ptolemy), or Lambert projections (with circular scales) or Mercator projection (with straight scales). There does not exist a conformal map such that all the ∞^2 great circles are mapped into straight lines of the plane.

The only map with straight scales is the Mercator projection whereas the only maps with circular scales are the stereographic and Lambert maps. Mercator's cartogram straightens out the loxodromes. This is possible not only for the sphere but also for any surface of revolution. We have shown that no further extension is possible.

[1] Kasner and De Cicco, "Ovals Should Be Used to Map Airplane Ranges," *Science News Letter*, 200 (March 25, 1944).

[2] Kasner and De Cicco, "Scale Curves in Conformal Maps," these PROCEEDINGS, **30**, 162–164 (1944).

[3] Kasner and De Cicco, "Scale Curves in General Cartography," *Ibid.*, **30**, 211–215 (1944).

[4] Kasner and De Cicco, "Geometry of Scale Curves in Conformal Maps," *Am. Jour. Math.*, **67**, 157–166 (1945).

[5] De Cicco, "Conformal Maps of Surfaces with Isothermal Systems of Scale Curves," *Ibid.*, **67** (1945).

[6] In our preceding paper, "The Laplace Equation in Space," published in the August PROCEEDINGS, theorems 4 and 5 should be ammended by replacing the Liouville inversion group by the similitude group. Thus while inversion in the plane preserves isothermal families of curves, inversion in space does not preserve isothermal families of surfaces. Conformality is necessary but not sufficient in three (or more) dimensions. This apparently has not been noticed before.

[7] Santola, "Un Theorema sobra Representacion Conforme," *Boletin Matematica Rosario*, 29–40 (1945). Many of the results of this interesting paper were given by Kasner in previous papers on differential elements (1906) and the invarient theory of horn angles (1912–1943), and in the papers by Kasner and De Cicco on cartography.